Gestion de projet & Produit digital

M2 SEP Année universitaire 2025-2026

FPI : FRANCE PROPERTY INSIGHT

Réalisé par :

Daniel **PHAN**
Nicolas **COLLIN**
Kim Ngan **THAI**
Perle **NDAYIZEYE**
Claudy **LINCY**

Sous la direction de :

Morgan **COUSIN**
Amor **KEZIOU**

Table des matières

Table des figures

Remerciements

Ce projet n'aurait pas été possible sans l'encadrement prodigué par le Master SEP : Statistique pour l'Évaluation et la Prévision de l'Université de Reims Champagne-Ardenne.

Nous remercions nos professeurs et les responsables du Master Emmanuelle GAUTHERAT, Jules MAES, et Philippe REGNAULT; mais en particulier Morgan COUSIN, responsable du module Gestion de Projet/Produit Digital, Amor KEZIOU, Frédéric BLANCHARD, et Philippe REGNAULT pour leurs enseignements, conseils et avis autour du Machine Learning.

Résumé

France Property Insight (FPI) est un produit d'analyses statistiques et de prédictions sur les **transactions immobilières en France**. Il permettrait aux **futurs acheteurs** d'estimer le budget nécessaire à leurs projets ou aux **propriétaires** de jauger la valeur de leurs biens sur plusieurs années.

Le produit sera développé sur cinq sprints de deux semaines, chacun donnant lieu à un rapport, un livrable fonctionnel et une démonstration auprès de Monsieur Morgan COUSIN, responsable du module Gestion de Projet Digitaux dans le cadre du Master 2 SEP : Statistique pour l'Évaluation et la Prévision.

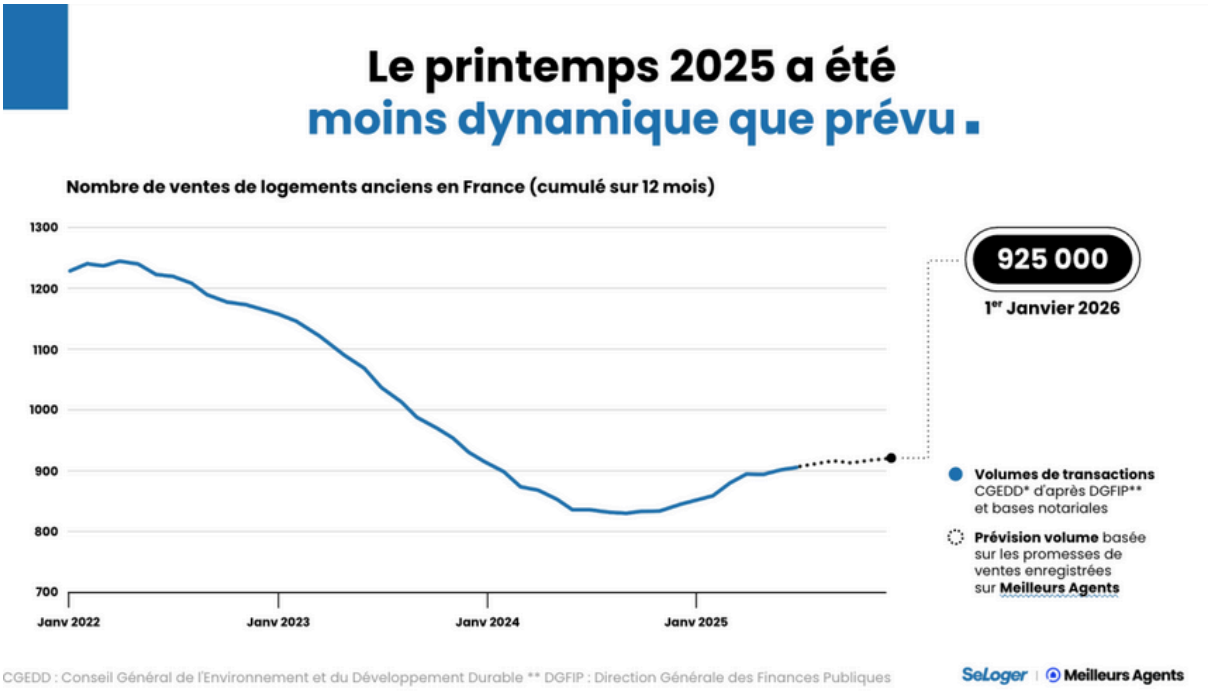
Nos modèles reposent sur le Machine Learning et s'appuient sur le cours d'Apprentissage Supervisé de Monsieur Amor KEZIOU, Maître de conférences à l'Université de Reims Champagne-Ardenne.

Le jeu de données « Demandes de valeurs foncières », publié et produit par la Direction Générale des Finances Publiques (DGFP), permet de connaître les transactions immobilières intervenues au cours des cinq dernières années sur le territoire métropolitain et les DOM-TOM, à l'exception de l'Alsace, de la Moselle et de Mayotte. Les informations proviennent des actes notariés et des données cadastrales.

Contexte

Dans le contexte actuel du marché immobilier français, **particuliers et investisseurs** sont confrontés à un défi difficile : déterminer le bon moment pour acheter ou vendre un bien, ainsi que son prix réaliste.

Figure 1 - Plus de 900 000 ventes de logements anciens sur les 12 derniers mois



Source : https://edito.meilleursagents.com

Après la détente monétaire amorcée dès 2024, beaucoup espéraient un redémarrage net de l'activité immobilière. Si le rebond a bien eu lieu en début d'année, il n'a pas tenu ses promesses sur la durée. Néanmoins, le nombre de transactions reste conséquent et l'État souhaite le voir remonter afin de retrouver les niveaux d'avant-crise.

Au-delà de la politique monétaire, plusieurs facteurs économiques et structurels influencent le marché en 2025. L'inflation, bien que sous contrôle ces derniers mois, reste un enjeu : par exemple, la Banque de France prévoyait un taux autour de 1 % pour 2025.

Les conditions de crédit se sont nettement améliorées : les taux moyens des prêts immobiliers s'établissaient à environ 3,24 % sur 20 ans et 3,32 % sur 25 ans début 2025.

Certaines expertises anticipaient même des taux pouvant descendre sous la barre des 3 % pour les meilleurs profils.

Pour mars 2025, le taux moyen national était autour de 3,45 % selon les régions. Cette amélioration de l'accès au crédit, combinée à une production de prêts en forte croissance (+50 % ou plus sur certains mois).

Parallèlement, les normes environnementales et la fiscalité modifient les comportements : l'obligation de rénovation pour certains logements, les dispositifs incitatifs ou restrictifs, participent à orienter les choix vers des biens mieux isolés ou adaptés aux nouveaux usages. Les dynamiques démographiques et sociétales entrent aussi en jeu : télétravail, mobilité accrue, souhait de logements plus spacieux (surtout en périphérie) transforment la demande.

Enfin, la géographie reste un facteur clé : les grandes métropoles continuent de connaître une forte pression prix et une demande soutenue, tandis que certaines zones rurales ou moyennes montrent des signes de stabilisation, voire de légère baisse des volumes et des prix. Dans ce contexte très segmenté, l'analyse locale et fine s'impose pour optimiser les décisions d'achat, de vente ou d'investissement.

FPI (France Property Insight) répond à ce besoin en proposant un outil analytique et prédictif des transactions immobilières. Il permet aux utilisateurs de visualiser les tendances du marché, d'estimer la valeur probable d'un bien selon ses caractéristiques et sa localisation, et d'identifier les zones ou types de biens les plus pertinents selon leurs objectifs. Dans un marché encore dynamique mais incertain, disposer de données fiables et d'analyses fines devient un atout stratégique pour orienter les décisions, limiter les risques et anticiper les évolutions futures, que ce soit pour un investissement locatif, l'acquisition d'une résidence principale ou la revente d'un patrimoine immobilier.

Source : https://edito.meilleursagents.com

Présentation des données

Le jeu de données **« Demandes de valeurs foncières »**, publié et produit par la Direction Générale des Finances Publiques (DGFiP), permet de connaître les transactions immobilières intervenues au cours des cinq dernières années sur le territoire métropolitain et les DOM-TOM, à l'exception de l'Alsace, de la Moselle et de Mayotte.
Les données contenues sont issues des actes notariés et des informations cadastrales.

Lien vers le jeu de données :
https://www.data.gouv.fr/datasets/demandes-de-valeurs-foncieres/

L'origine des données

Les informations mises à disposition sont issues du traitement informatisé « Demande de valeurs foncières » alimenté par la **« Base nationale des données patrimoniales »** (BNDP).

L'application BNDP, qui recense les données patrimoniales contenues dans les documents déposés par les redevables ou leurs représentants dans les services en charge de la publicité foncière et de l'enregistrement, est alimentée par les traitements informatisés de l'administration fiscale relatifs à la documentation cadastrale (traitement « Majic ») et à la publicité foncière (traitement « Fidji »).

Les informations figurant dans chaque fichier sont donc **issues du système d'information de la DGFiP** (Direction Générale des Finances Publiques) , après publication des actes par le service de la publicité foncière, et complément des éléments cadastraux (référence cadastrale, nature des biens et descriptif des biens).

Important : le descriptif des biens de l'acte notarié n'est pas repris dans le fichier, à l'exception de la surface Carrez lorsque celle-ci est mentionnée.
NB : le contenu des fichiers dépend donc des informations qui auront été dûment publiées par le service de la publicité foncière.

Le périmètre géographique

Les informations diffusées sont issues des mutations publiées dans les services de la publicité foncière de **l'ensemble du territoire de la France métropolitaine, à l'exception des départements du Bas-Rhin, du Haut-Rhin et de Moselle**; les données les concernant relèvent du « livre foncier » et ne sont pas mobilisées dans la base nationale des documents patrimoniaux (BNDP), ainsi que des départements et régions d'outre-mer, excepté Mayotte.

La mise à jour des données

En application du décret du 28 décembre 2018, les informations communiquées font l'objet d'une mise à jour semestrielle.

Chaque année, une première diffusion sera effectuée en avril, présentant les mutations intervenues au cours des cinq dernières années et ayant fait l'objet d'une publication par un service de publicité foncière avant le 31 décembre de l'année précédente. La diffusion du mois d'avril concerne en conséquence cinq millésimes soit 10 semestres. Une seconde diffusion sera effectuée en octobre portant sur les mutations intervenues au cours des cinq dernières années et ayant fait l'objet d'une publication par un service de publicité foncière avant le 30 juin de l'année en cours.

L'attention est appelée sur le fait qu'en avril comme en octobre, compte tenu des publications effectuées au cours du dernier semestre pouvant porter sur des mutations intervenues lors de semestres précédents, l'ensemble des fichiers annuels sont actualisés.

Chaque mise à jour supprime puis remplace la totalité des fichiers précédemment publiés.
La dernière mise à jour date du 6 avril 2025.

Dictionnaire des variables

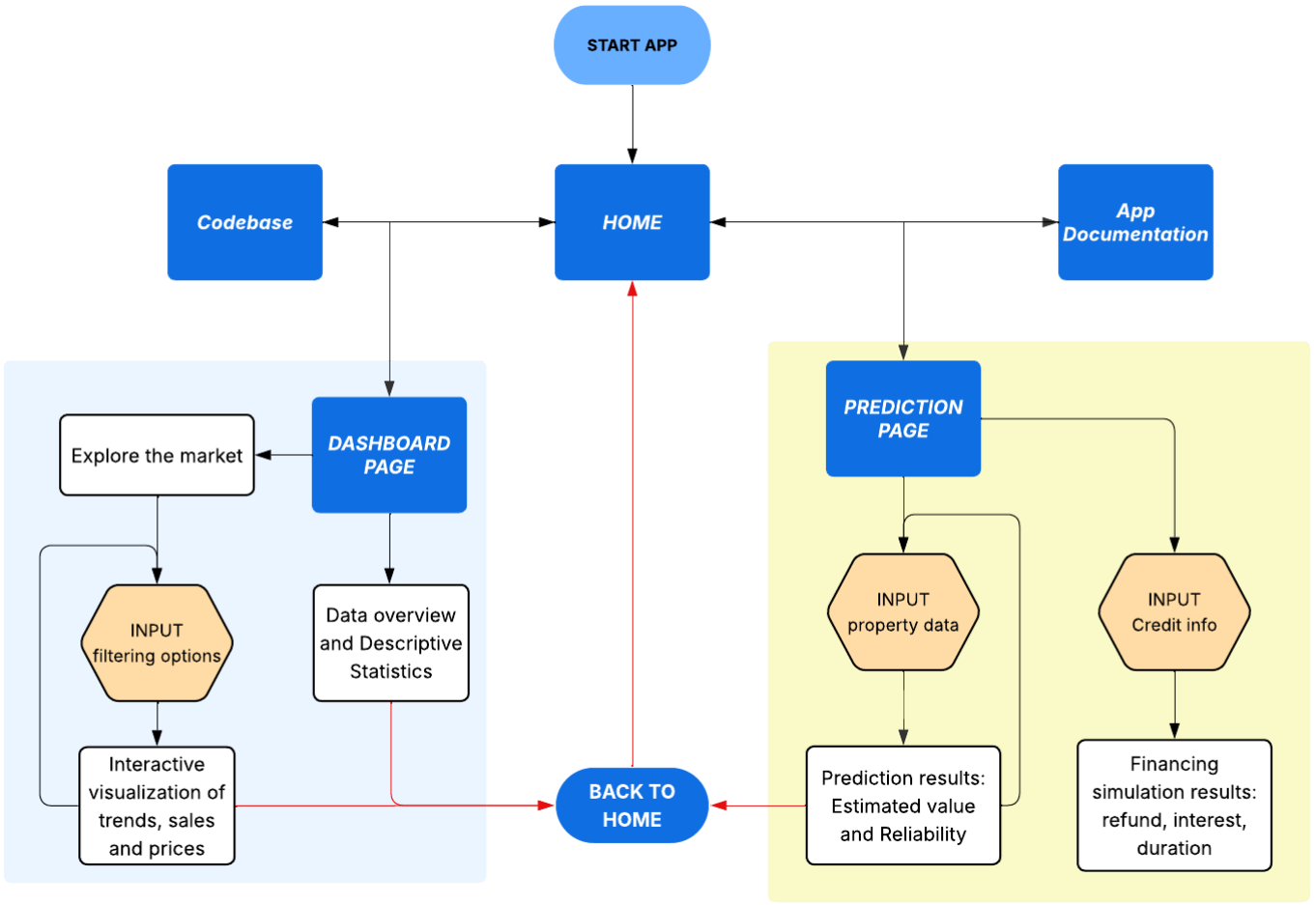
Le jeu de données brut propose 43 variables, cependant nous allons retenir les quelques variables suivantes (voir les raisons de rejets ici : Variables non retenues et raisons de rejet :)

Libellé originel	Descriptif	Nouveau nom dans notre base de données
Valeur foncière	Montant de l'évaluation déclaré(e) dans le cadre d'une mutation à titre onéreux. Inclut : frais d'agence si à charge du vendeur et TVA Exclut : frais d'agence si à charge de l'acquéreur, frais de notaires, valeur des biens meubles stipulée dans l'acte de mutation	property_value
Date de mutation	Format JJ/MM/AAAA	transaction_date
Nature de la mutation	Vente, vente en l'état futur d'achèvement, vente de terrain à bâtir, adjudication, expropriation ou échange	transaction_type
Code postal	Référence cadastrale de la parcelle.	postal_code
Code département	Référence cadastrale de la parcelle.	department_code
Code commune	Référence cadastrale de la parcelle.	town_code
Commune	-	town
Code type local	1 : maison, 2 : appartement, 3 : dépendance (isolée), 4 : local industriel et commercial ou assimilés	property_type_code
Type local	-	property_type
Surface réelle bâtie	La surface réelle est la surface mesurée au sol entre les murs ou séparations et arrondie au mètre carré inférieur. Les surfaces des dépendances ne sont pas prises en compte.	building_area
Nombre pièces principales	Les cuisines, salles d'eau et dépendances ne sont pas prises en compte.	main_rooms
Surface terrain	Contenance du terrain	land_area

Présentation de l'application

Fonctionnalités principales

Figure 2 - Logigramme de FPI



Source : les auteurs, 2025

Logigramme de l’application FPI

Cette structure illustre le parcours utilisateur au sein de notre application, qui s'articule autour de trois fonctions clés accessibles depuis la page **Accueil** (HOME) :

1. **Tableau de bord (DASHBOARD PAGE) :** cette section est dédiée à l'analyse du marché immobilier. Les utilisateurs peuvent explorer les tendances, les ventes et les prix grâce à des visualisations interactives.
2. **Prédiction (PREDICTION PAGE) :** cette section permet d'effectuer deux types de simulations :
    - Estimation de la valeur d'une propriété.
    - Simulation des conditions de financement (remboursement, intérêts, durée).
3. **Documentation (CODEBASE et API DOC) :** accès aux ressources d'information sur l'application.

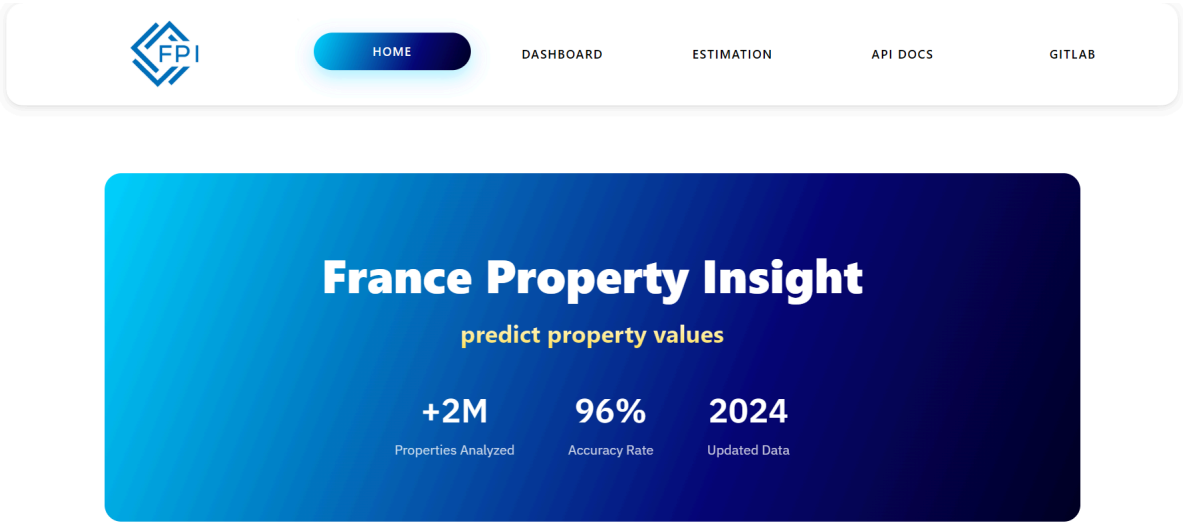
L'utilisateur peut, à tout moment après avoir utilisé les fonctionnalités principales, revenir facilement à l'Accueil ou naviguer vers une autre section.

Interface utilisateur

L'application, développée avec Gradio, est conçue pour être simple, intuitive et interactive. Elle se compose de **trois pages principales** : *Accueil*, *Dashboard* et *Prédiction*.

1. **Page d'accueil**

Figure 3 - page d'accueil



Source : FPI, 2025

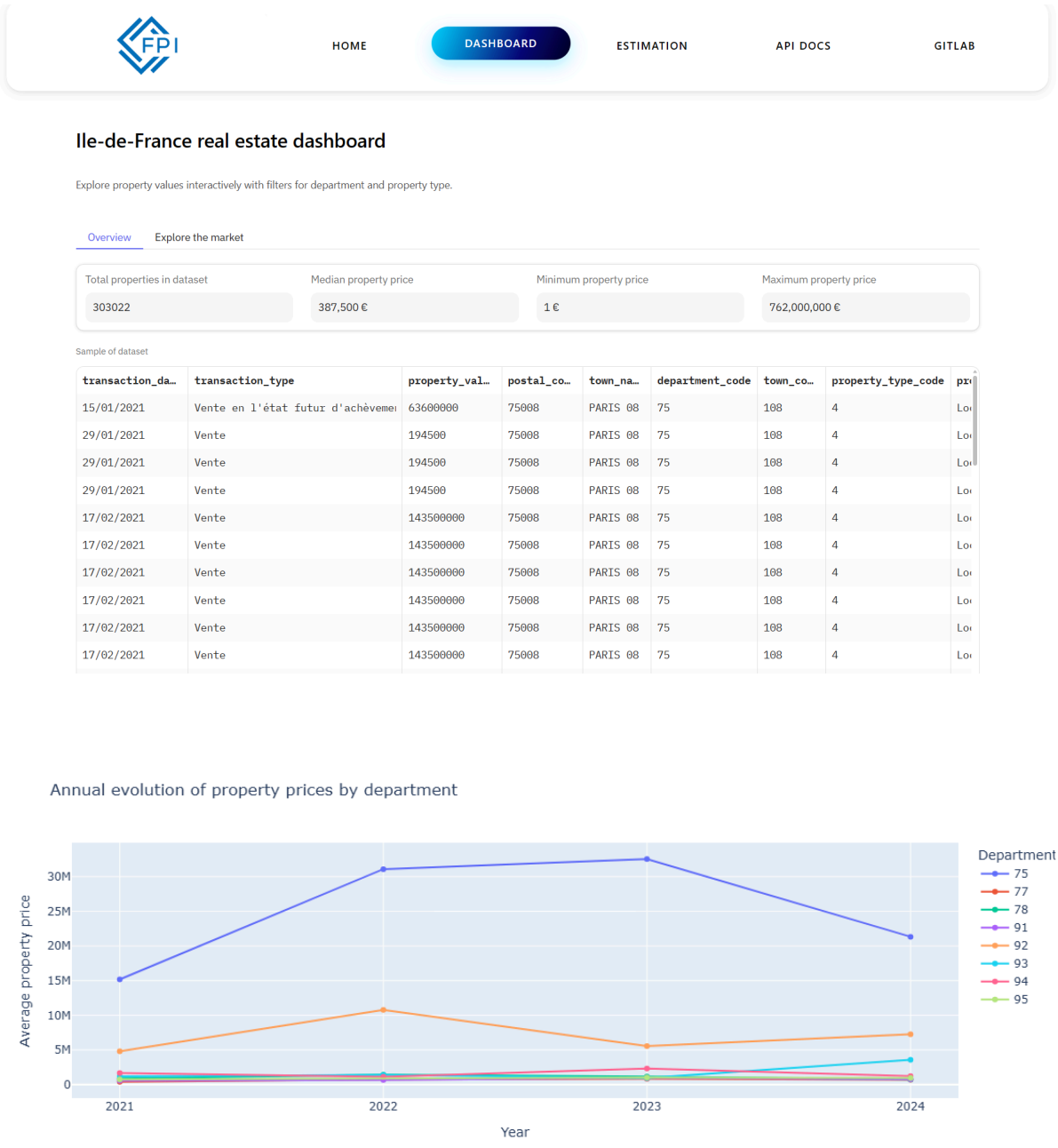
Cette page constitue le point d'entrée vers les principales fonctionnalités de l'application : le tableau de bord, l'outil de prédiction, ainsi que l'accès à la documentation (API Docs) et au dépôt GitLab du projet.

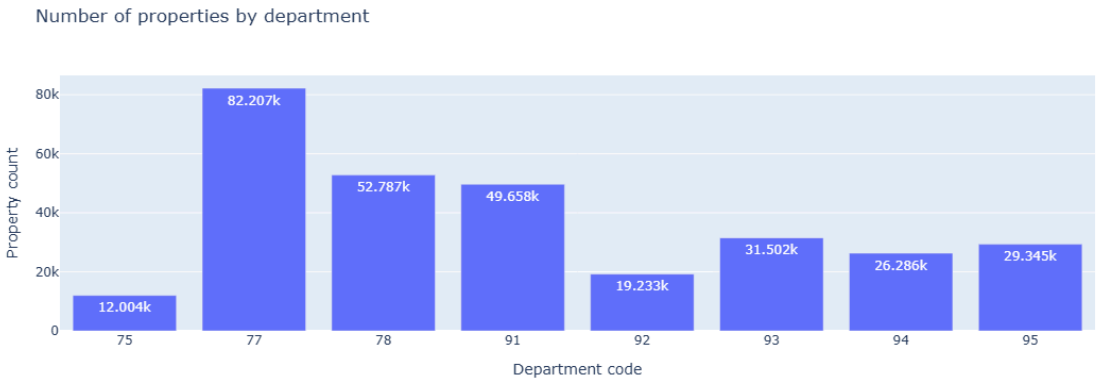
2. Page du Dashboard

Le tableau de bord est l'outil d'analyse du marché de l'application. Il offre une vue d'ensemble des données pour guider l'utilisateur dans l'exploration du marché.

a. Tab Overview

Figure 4 - page du tableau de bord (Overview)



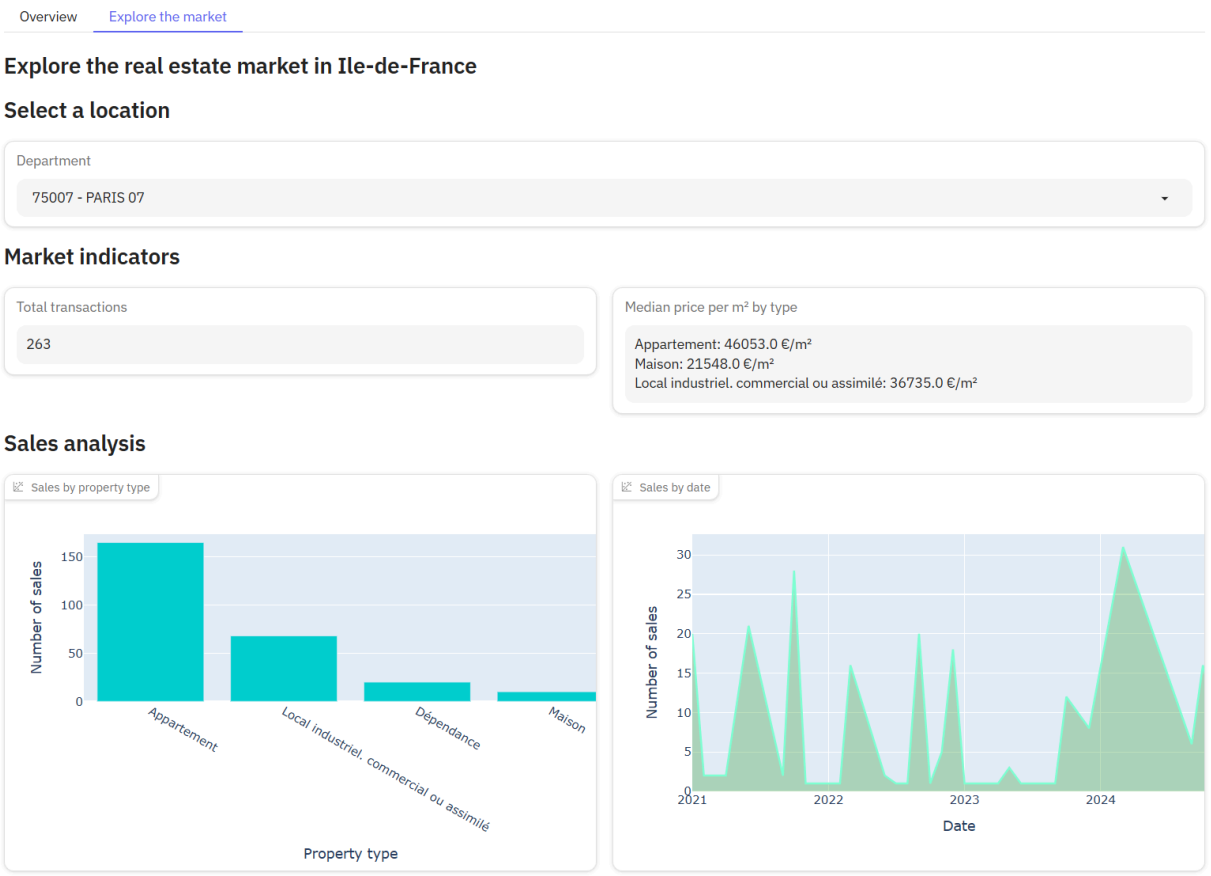


Source : FPI, 2025

Overview regroupe des statistiques descriptives ainsi qu'un exemple de notre base. L'utilisateur peut interagir avec les visualisations pour explorer les données selon différents critères et mieux comprendre les tendances du marché immobilier.

b. Tab Explore the market

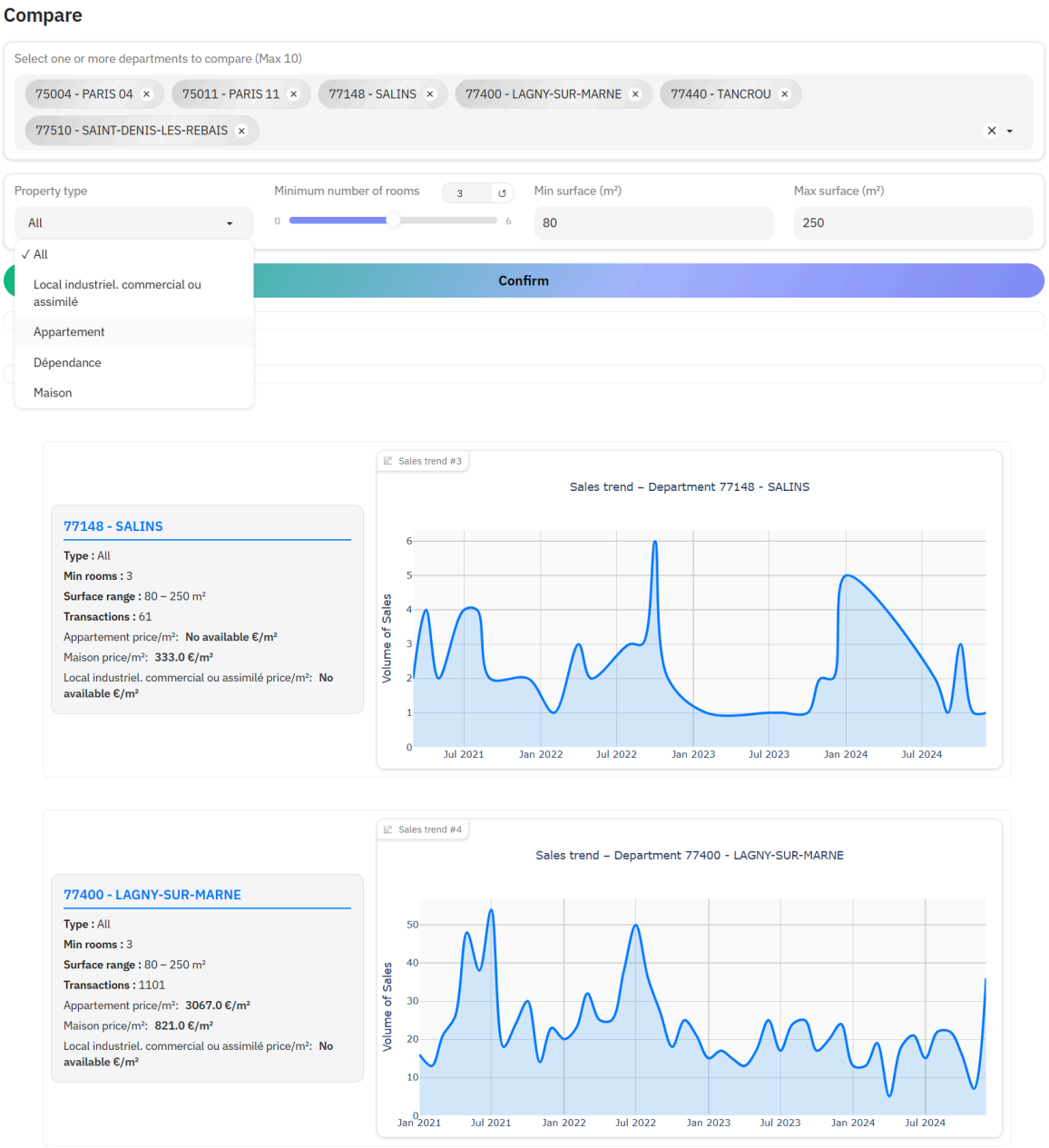
Figure 5 - page Explore the market



Source : FPI, 2025

L'utilisateur peut sélectionner une zone à explorer depuis une liste déroulante. La page affiche ensuite une étude de marché avec les principaux indicateurs, tels que le nombre de biens vendus ou le prix médian au mètre carré selon le type de bien. Des graphiques interactifs présentent également la répartition des ventes par type de bien et par année, permettant à l'utilisateur d'explorer les données en détail.

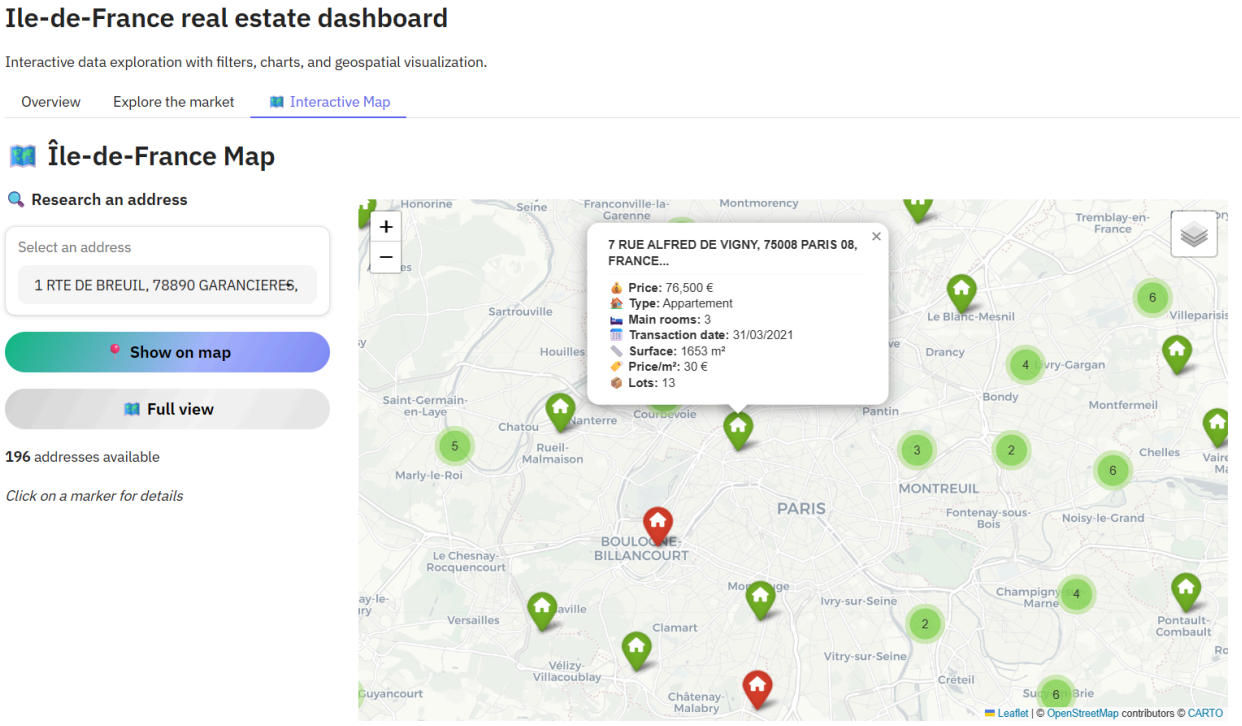
Figure 6 - comparer les tendances



Source : FPI, 2025

Par ailleurs, l'utilisateur peut comparer les tendances entre plusieurs villes ou département avec les filtres spécifiques tels que le type de bien, minimum principale pièces, l'intervalle de la superficie qu'il veut savoir. Il affiche ensuite les informations sur le nombre de ventes, le prix par m² ainsi que l'évolution des prix.

Figure 7 - carte interactive



Source : FPI, 2025

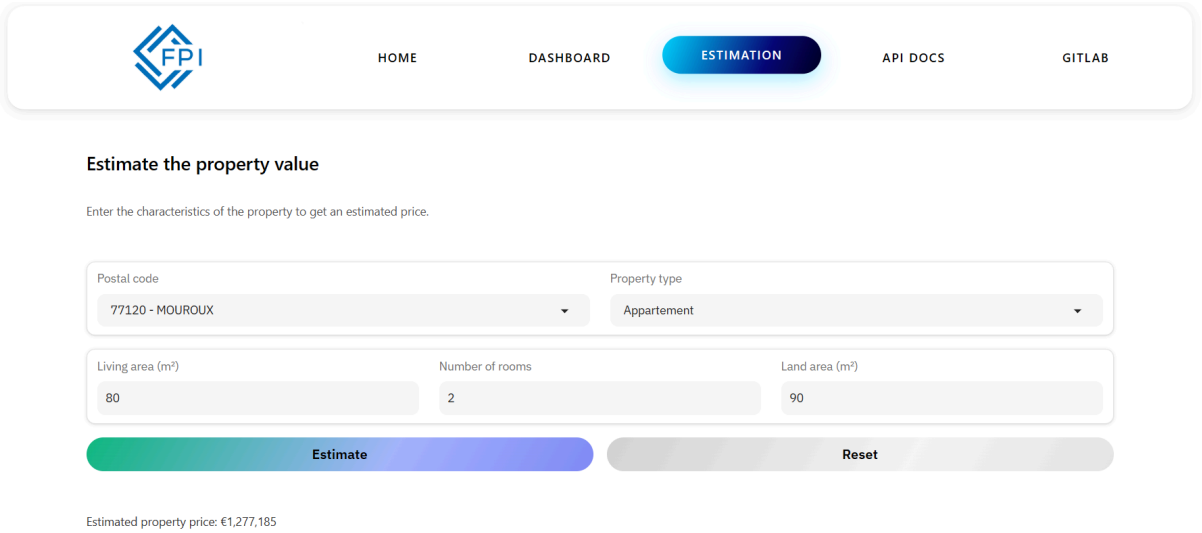
La dernière section du tableau de bord présente une carte interactive des transactions immobilières en Île-de-France.

Les biens sont représentés sous forme de points géolocalisés à partir de leur adresse. L'utilisateur peut rechercher une adresse via la barre de recherche située à gauche et interagir avec les points affichés sur la carte afin de consulter les informations associées à chaque bien, notamment le prix au mètre carré et le nombre de lots.

Cette visualisation permet d'analyser la répartition spatiale des transactions immobilières et d'identifier rapidement les zones de concentration ou de disparité des prix au sein de la région.

3. **Page Prédiction**

Figure 8 - page de prédiction



Source : FPI, 2025

Cette page permet à l'utilisateur d'effectuer des prédictions personnalisées à partir de variables qu'il peut saisir et sélectionner.

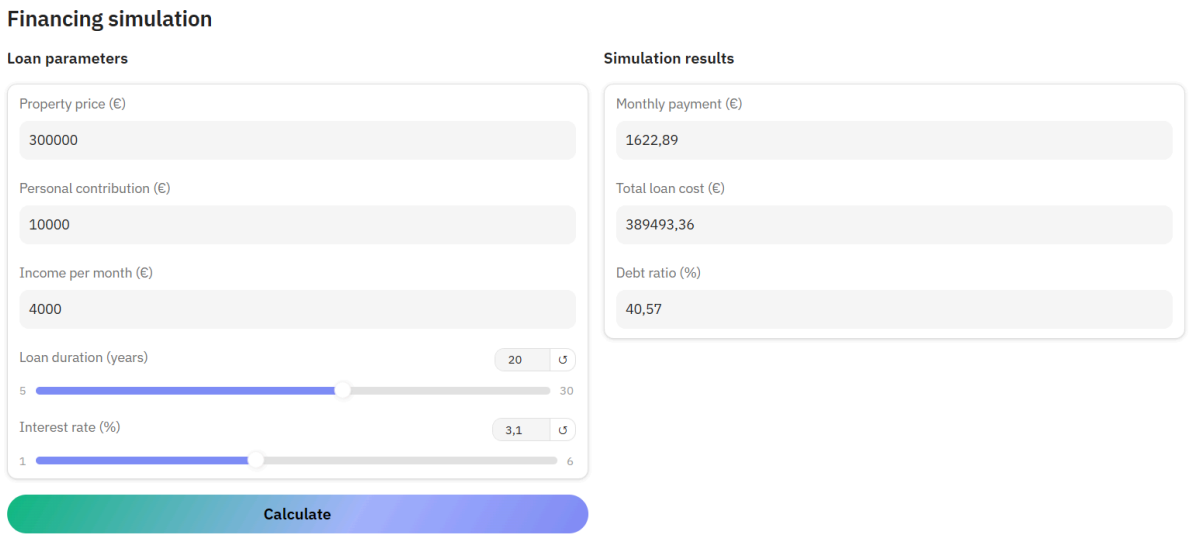
L'utilisateur y renseigne notamment les caractéristiques de son bien tels que :

- La localisation (code postale, ville, quartier).
- Le type de bien (appartement, maison).
- La superficie.
- Le nombre de pièces.

Le résultat est généré instantanément après la saisie. Cette page est également utilisée pour la simulation de financement.

Financing simulation

Figure 9 - page de prédiction : financing simulation

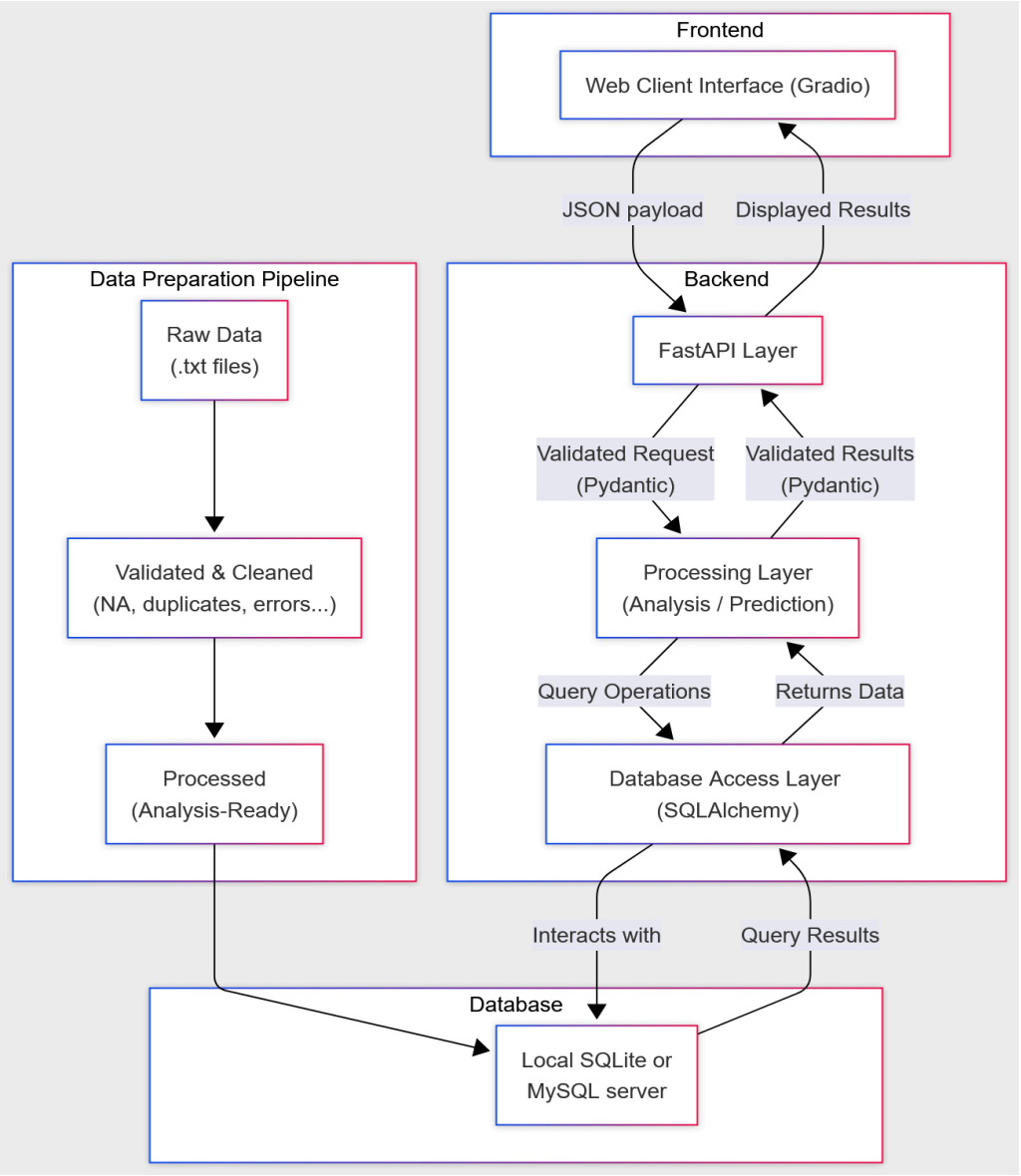


Source : FPI, 2025

L'utilisateur peut également effectuer une simulation de crédit en indiquant le montant souhaité, son apport personnel, son revenu ainsi que la durée et le taux d'intérêt. L'outil calcule ensuite le montant des mensualités, le coût total du crédit et le ratio d'endettement.

Flux des données

Figure 10 - dataflow



Source : les auteurs, 2025

Nous prévoyons une **architecture de données en médaillons** (Bronze/Argent/Or) ou données brutes/nettoyées/préparées (=prêtes pour analyse/modélisation).

Du côté utilisateur, ses requêtes seront envoyées sous forme de payload JSON, transmises à travers une API web développée avec FastAPI, puis validées par Pydantic avant d'être traitées par la couche applicative. Cette couche calculatoire accèdera aux données via SQLAlchemy, afin de faciliter la **portabilité entre différents systèmes de gestion de base de données** tels que SQLite, MySQL ou PostgreSQL.

Nettoyage et préparation des données (NA, dups, outliers)

Concernant la préparation des données, nous avons créé une fonction qui fusionne les tables en .txt, elle extrait ensuite les variables intéressantes à explorer et enrichir notre produit, tout en renommant ces dernières. Après avoir choisi les variables qui nous intéressent, nous supprimons les NA et les doublons ainsi que les valeurs aberrantes.

Nous avons également créé des modèles de données Pydantic pour la validation de données, forcant ainsi chaque variable à respecter des contraintes de valeur, types, et format. Et formalisant ainsi la structure attendue de chaque ligne dans chaque fichier de données.

Les données complètes sur toute la France étant trop lourdes pour un dépôt GitLab, nous avons décidé de se concentrer temporairement sur la région d'Île de France, elle comprend 8 départements : Paris (75), Seine-et-Marne (77), Yvelines (78), Essonne (91), Hauts-de-Seine (92), Seine-Saint-Denis (93), Val-de-Marne (94) et Val-d'Oise (95). Nous avons appliqué un filtre afin de ne garder que les transactions dans les départements de la région en utilisant la variable “department_code” (=code département).

Toutes les fonctionnalités de FPI seront toutefois scalables sur les données de toute la France, tant que l’utilisateur est prêt à télécharger localement toutes les données (et installer FPI via les instructions présentes sur notre GitLab).

Effectif des variables manquantes

Variable	Value
Identifiant_de_document	1979256
Reference_document	1979256
1_Articles_CGI	1979256
2_Articles_CGI	1979256
3_Articles_CGI	1979256
4_Articles_CGI	1979256
5_Articles_CGI	1979256
No_disposition	0
Date_mutation	0
Nature_mutation	0
Valeur_fonciere	16722
No_voie	252995
B/T/Q	1863417
Type_de_voie	197191
Code_voie	30670
Voie	30730
Code_postal	30694
Commune	0
Code_departement	0
Code_commune	0
Prefixe_de_section	1971282
Section	0
No_plan	0
No_Volume	1971133
1er_lot	661458
Surface_Carrez_du_1er_lot	1540861
2eme_lot	1529381
Surface_Carrez_du_2eme_lot	1847356
3eme_lot	1916420
Surface_Carrez_du_3eme_lot	1967306
4eme_lot	1957490
Surface_Carrez_du_4eme_lot	1976372
5eme_lot	1970109
Surface_Carrez_du_5eme_lot	1978256
Nombre_de_lots	0
Code_type_local	336625
Type_local	336625
Identifiant_local	1979256
Surface_reelle_bati	337530
Nombre_pieces_principales	337530
Nature_culture	1326517
Nature_culture_speciale	1917987
Surface_terrain	1326517

Effectif des outliers

Variable	Value
Identifiant_de_document	0
Reference_document	0
1_Articles_CGI	0
2_Articles_CGI	0
3_Articles_CGI	0
4_Articles_CGI	0
5_Articles_CGI	0
No_disposition	50776
Valeur_fonciere	282163
Code_departement	0
Code_commune	27236
No_plan	158528
Surface_Carrez_du_1er_lot	16376
Surface_Carrez_du_2eme_lot	4417
Surface_Carrez_du_3eme_lot	699
Surface_Carrez_du_4eme_lot	209
Surface_Carrez_du_5eme_lot	78
Nombre_de_lots	62836
Identifiant_local	0

Effectif du type de propriété

Variable	Value
Dépendance	788203
Appartement	557081
nan	336625
Maison	210842
Local industriel. commercial ou assimilé	86505

Variables non retenues et raisons de rejet :

Le nombre de valeurs manquantes a été calculé plus tôt dans cette partie :

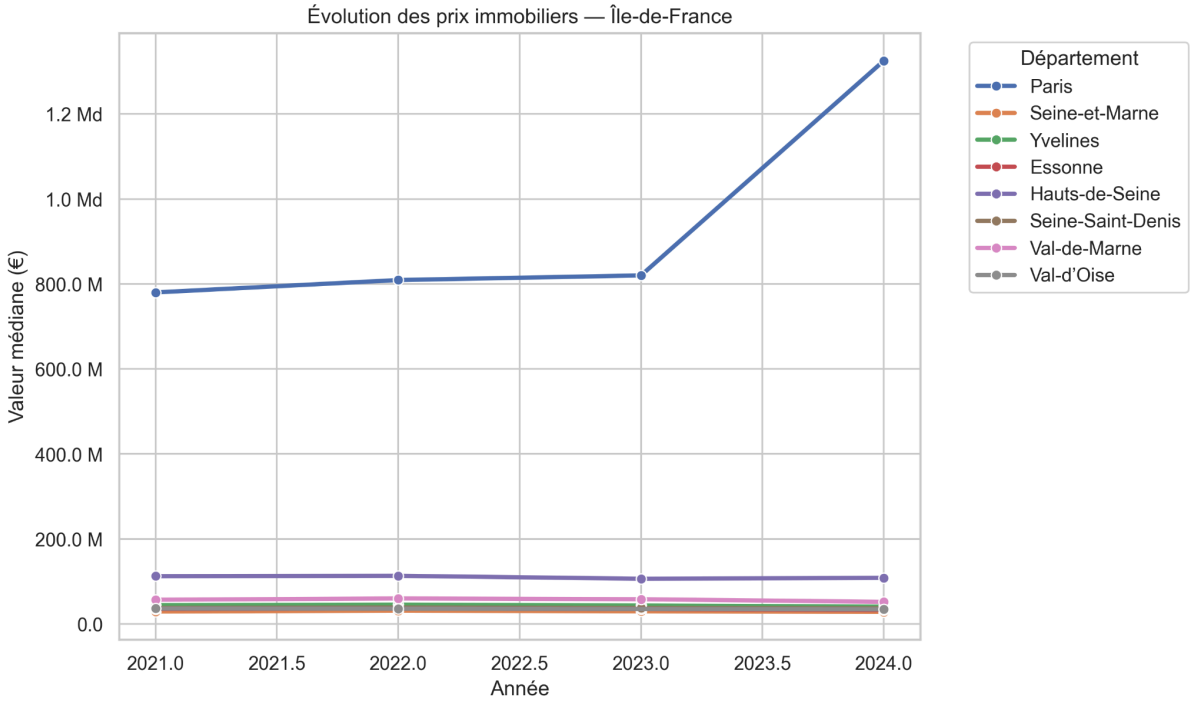
Nettoyage et préparation des données (NA, dups, outliers)

Libellé	Raison de rejet
Code service CH	Non restitué
Référence document	Non restitué
Article CGI 1-5	Non restitué
N° de voie/B/T/Q	Valeurs manquantes
Type de voie/Code voie/Voie	Valeurs manquantes
Préfixe de section/Section	Valeurs manquantes
N° de plan/N° de volume	Valeurs manquantes
1er-5ème lot	Valeurs manquantes
Surface Carrez du 1er-5ème lot	Valeurs manquantes
Identifiant local	Non restitué
Code nature culture	Valeurs manquantes
Nature culture spéciale	Valeurs manquantes

Analyse des données

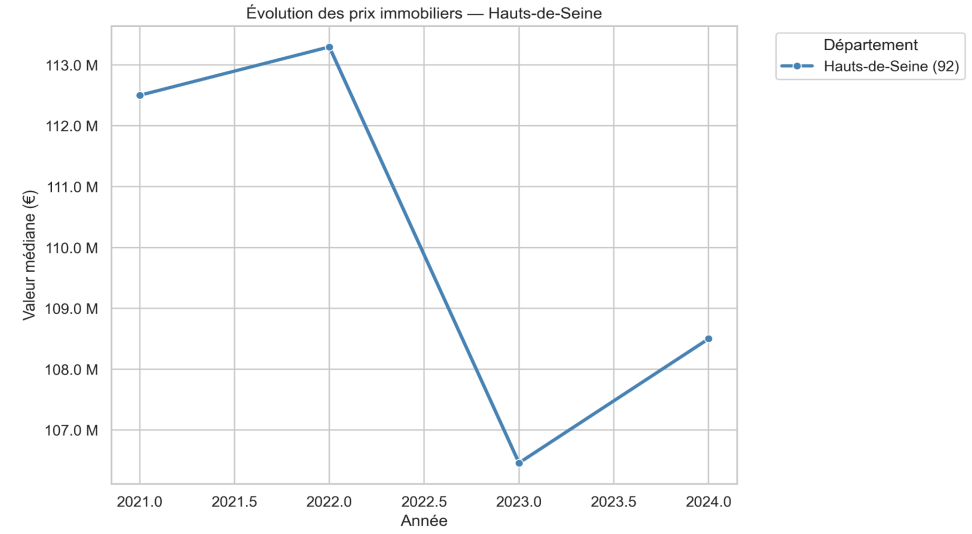
Analyse exploratoire: Graphes des tendances

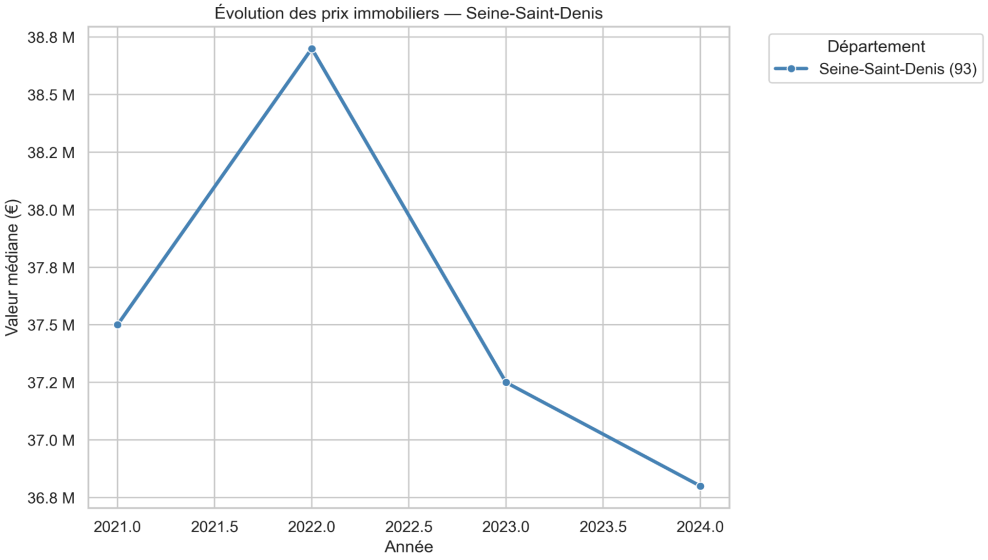
Figure 11 - évolution des prix immobiliers en Île-de-France



Une première lecture du graphique des tendances révèle une forte disparité entre Paris et les autres départements de la région. Cette différence s'explique par le niveau particulièrement élevé des prix immobiliers dans la capitale, nettement supérieur à celui observé ailleurs en Île-de-France.

L’observation des graphiques d’évolution, département par département, met en évidence les spécificités locales du marché:





Les Hauts-de-Seine se positionnent comme le deuxième département le plus cher en matière de prix immobilier, tandis que la Seine-Saint-Denis affiche les tarifs les plus bas de la région. Entre 2021 et 2022, les deux départements ont connu une hausse des prix, marquée par un pic particulièrement prononcé en Seine-Saint-Denis. Toutefois, cette tendance s’est inversée dès l’année suivante, avec une baisse observable des prix.

Modélisation et prédictions

Sélection et préparation des variables

L'analyse de la valeur foncière repose sur un ensemble de variables clés issues des données de transaction de 2021. Le modèle a été initialement entraîné sur les données de l'année 2021, qui servent de référence pour l'apprentissage des relations entre les caractéristiques des biens et leur prix.

Après nettoyage et sélection des variables exploitables, la modélisation se base sur les caractéristiques fondamentales du bien et sa localisation.

Catégorie	**Variables clés**	**Rôle dans le modèle**
Caractéristiques du bien	`land_area`, `building_area`, `property_type`, `main_rooms`	Décrivent la taille et la nature physique du bien.
Localisation	`department_code`, `postal_code`, `town`	Capturent l'effet "micro-marché" et les disparités territoriales.

Les variables ont été transformées pour n'utiliser que les transaction de type ventes et la découpe des données a été effectuée à 70% pour l'apprentissage et 30% pour la validation.

Évaluation et comparaison des modèles

Plusieurs modèles de régression ont été évalués. Une première comparaison visait à valider l'intuition selon laquelle la complexité du marché immobilier français, caractérisé par des interactions non-linéaires et de fortes disparités, nécessite des modèles d'ensemble (*ensemble models*).

Modèle	MAE (€)	RMSE (€)	R2	Famille
Random Forest	$3{,}75 \times 10^7$	$2{,}96 \times 10^8$	0,885	Ensemble
Gradient Boosting	$8{,}80 \times 10^7$	$3{,}79 \times 10^8$	0,811	Boosting
Decision Tree	$3{,}78 \times 10^7$	81×10^8	0,809	Base
Lasso	$1{,}56 \times 10^8$	$7{,}21 \times 10^8$	0,316	Linéaire
Ridge	$1{,}56 \times 10^8$	$7{,}22 \times 10^8$	0,315	Linéaire

Conclusion du choix :

Le Random Forest Regressor s'impose pour l'instant comme le modèle le plus performant. Avec un R^2 de 0,885, il explique près de 89% de la variance totale des prix immobiliers, surpassant largement les modèles linéaires (R^2 approx 0,32). Cette performance confirme que les relations entre les caractéristiques et le prix sont fortement non-linéaires.

Le Random Forest est retenu comme modèle de référence pour cette version en raison de:

- Sa précision supérieure (MAE minimal).
- Sa robustesse face aux valeurs extrêmes et à l'hétérogénéité des données.
- Sa capacité à fournir une interprétation claire de l'importance des variables.
- une stabilité opérationnelle lors de mises en production.

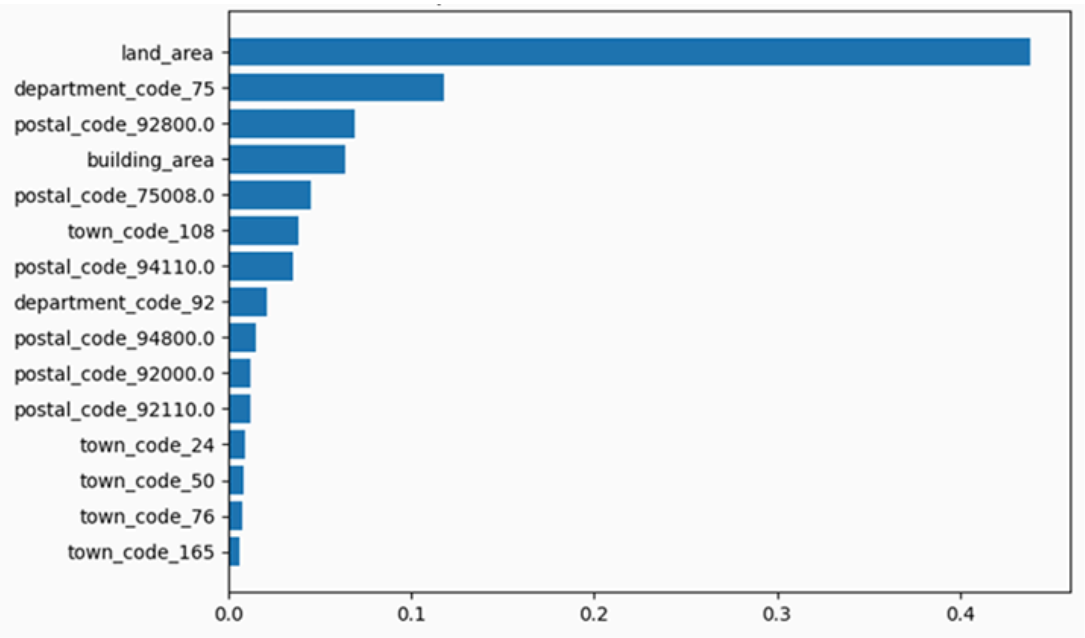
Interprétabilité et analyse des variables

L'analyse de l'importance des variables confirme les intuitions des professionnels de l'immobilier, mettant en évidence la primauté de la localisation et des surfaces.

Les facteurs les plus déterminants, dans l'ordre de leur contribution au modèle, sont:

- `Land area` (surface du terrain) : impact majeur, notamment dans les zones moins denses
- `Localisation` : rôle déterminant des départements 75, 92
- `Building area` (Surface habitable) et nombre de pièces
- `Typologie` (maison / appartement)

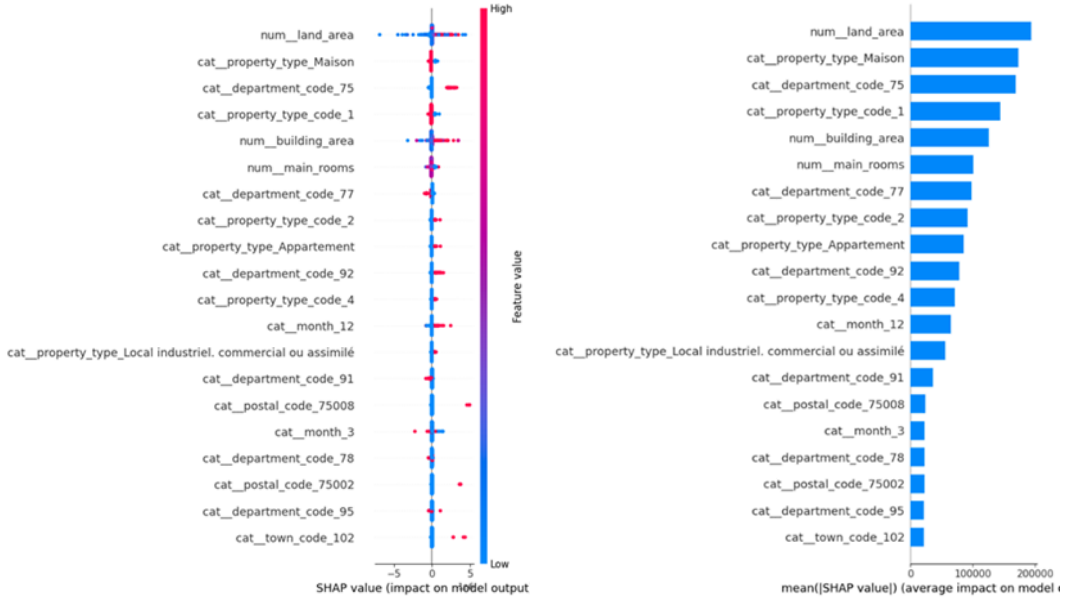
Figure 12 — Importance des variables (Random Forest)



Cette organisation confirme des priorités bien connues des agents immobiliers : localisation, typologie, et surface.

Analyse SHAP : compréhension fine du modèle

Graphiques : Importance SHAP (global et summary plot)



Les résultats montrent que :

- Les surfaces (terrain et habitable) sont les premiers moteurs d'augmentation du prix.
- La typologie Maison exerce un effet positif très marqué.
- Les effets départementaux révèlent l'existence de micro-marchés premium (Paris, Hauts-de-Seine).
- Les variables temporelles (mois, année) jouent un rôle secondaire mais cohérent avec la saisonnalité du marché.

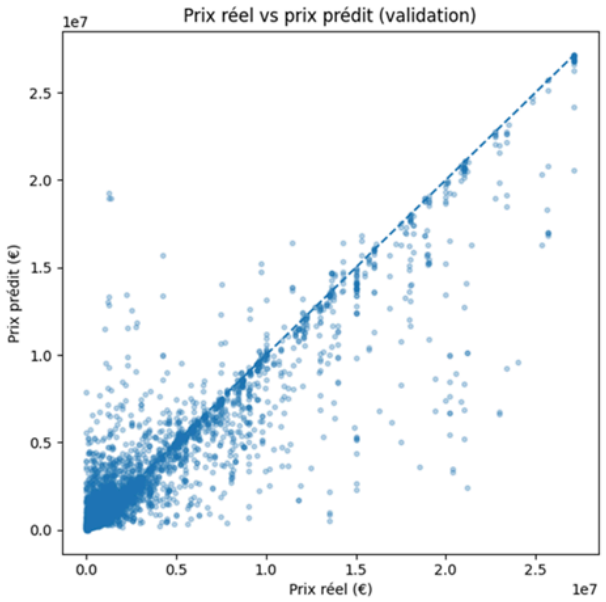
Ces apports permettent d'avoir une vision transparente du fonctionnement du modèle : les décisions ne sont pas arbitraires mais correspondent à des logiques économiques identifiables.

Analyse des erreurs

L'étude des résidus met en évidence une bonne stabilité globale du modèle, avec toutefois des difficultés sur :

- les biens très haut de gamme (> 1,5 M€), souvent sous-estimés ;
- certains micro-marchés premium (départements 75 et 92) ;
- les très petites surfaces, pour lesquelles l'erreur relative devient instable.

Figure 13 — Résidus & erreurs par département



Département	MAE (€)
77	117 736
95	158 657
91	187 314
78	215 910
93	285 879
94	315 896
92	653 525
75	1 617 344

Ces zones correspondent à des marchés atypiques, où la faible représentativité des transactions complique la modélisation.

Limites et perspectives

Le Random Forest Regressor fournit une solution robuste et précise pour la prédiction des prix immobiliers, expliquant près de 90% de la variance observée.

Cependant, plusieurs limites persistent :

- sous-estimation des biens premium ou atypiques ;
- absence de certaines variables clés (étage, rénovation, vue, ascenseur) ;
- forte dépendance aux micro-marchés locaux ;
- erreurs relatives élevées pour les très petites surfaces.

Déroulement des sprints

Sprint 1

Le projet a débuté par une réflexion sur les **objectifs individuels** de chaque membre, afin d'assigner les rôles et les tâches de manière cohérente avec les compétences et attentes de chacun.
La communication s'effectue principalement via Discord ou en présentiel, avec des réunions hebdomadaires et un suivi individuel assuré par le Scrum Master.

Ce premier sprint a été consacré à la **conception du produit et à la planification du projet**. L'équipe a identifié un besoin concret, collecté des idées de fonctionnalités, et élaboré une première version de la roadmap et un backlog (voir Annexe) en pensant à la **scalabilité** de l'application.

À l'issue de ce Sprint 1, nous considérons les fondations organisationnelles et techniques du projet satisfaisantes. Nous nous pencherons en priorité lors du Sprint 2 sur les analyses exploratoires des données, visualisations et modélisations.

Feedback Sprint 1

Points positifs :
-utilisation de données officielles open data (=data.gouv, INSEE…)

Points négatifs :
-nous aurions dû faire un premier modèle naïf pour avoir une vrai livrable utilisable
-manque de clarté sur la variable cible, penser à faire une slide uniquement pour elle
-revoir la présentation des données dans l'ensemble
-abandon des features login et historique (trop complexe et hors-sujet)
-attention à la valeur perçue lors d'une démo client

Conclusion : Se concentrer sur nos 2 pages principales : Dashboard et Prédiction.

Sprint 2

Ce deuxième sprint a eu lieu pendant la pause pédagogique de la Toussaint et le cumuls de plusieurs rendus à la rentrée, ce qui a été perturbant pour les disponibilités et l'organisation de réunions.
Nous avons d'abord commencé par renouveler notre roadmap et logigramme, et commencé dès le début à réfléchir pour la présentation, notamment autour de la valeur perçue et la visualisation des données.
Similairement au premier sprint, chaque membre a pu travailler individuellement sur sa branche sur des features découpées de manière à faciliter la fusion du travail de tous.

Le sprint 3 se concentrera principalement sur l'amélioration des fonctionnalités existantes en termes de performances ou d'options de visualisation.

Feedback Sprint 2

Points positifs :
-bonne documentation, présentation, et structure du git
-bonne définition des objectifs

Points négatifs :
-manque d'une bonne phrase d'accroche pour vendre le produit
-tests unitaires non harmonisés
-manque de détails sur le contexte économique
-présentation de la roadmap difficile
-bugs visuels sur l'application

Conclusion : Continuer à ajouter les features importantes, revoir la présentation du produit/contexte, harmoniser le code.

Sprint 3

Pour ce sprint 3, l'équipe s'est d'abord réunie très tôt pour faire une rétrospective des deux derniers sprints, notamment pour identifier des possibles améliorations dans la communication ou la collaboration.
Par manque de communication lors du sprint 2, les tâches assignées étaients trop vagues et beaucoup de code/travail s'est avéré inutile, en plus d'avoir été partagé (push) d'un seul coup et trop tard pour faire des review de codes ou encore nettoyer correctement le dépôt GitLab.
Nous en avons profité pour aussi faire le point sur toute notre chaîne du traitement de données, l'architecture en médaillon, optimisation du formulaire, et l'intégration de routes FastAPI et la transition des csv vers SQLite.

Sprint 4

Claudy notre Data Scientist a eu des problèmes de santé, et Daniel notre Product Owner a prévu des déplacements professionnels, le rendant peu disponible, notamment le jour du rendu et de la présentation client; ce qui nous a amené à revoir les objectifs à la baisse, pour rester réaliste.

L'accroche en présentation a été retravaillée, et nous continuons d'affiner les scénarios utilisateurs. Nous aborderons le sprint 5 dans le but de combler toutes les faiblesses du projet comme l'analyse des données, des fonctionnalités mineures et quelques défauts viuels de l'application.

Sprint 5

Pour ce dernier sprint, nous avons réussi à rendre l'application environ 10 fois plus rapide sur les prédictions grâce à la mise en place d'un serveur indépendant, rajouté une sécurité sur les saisies des prédictions, ajouté une carte interactive pour visualiser géographiquement les transactions de la base de données et ajusté certains éléments visuels.

Environnement de développement et choix techniques

Nous avons visé à construire un environnement de travail structuré et facilement reproductible et fait des choix techniques visant à développer une application scalable pour faciliter toute évolution du projet.

Dépôt **GitLab** :

- Assignation des rôles Gestionnaire et Développeurs (Maintainer/Developer)
- Protection de la branche main
- Workflow en Merge Requests
- Pipeline CI/CD pour automatiser les contrôles de qualité et les tests
- Architecture des données en médaillons et organisation du dépôt

Autres outils :

- Mermaid pour la documentation graphique
- pdoc pour une API documentation du code
- Python 3.13 sur Visual Studio Code
- Gestion des dépendances via uv
- Conteneurisation avec Docker (et .devcontainer)
- Développement de l'interface utilisateur avec Gradio
- Hébergement en ligne sur Render.com pour utilisation immédaite sans installation.
- Utilisation de SQLAlchemy pour assurer portabilité et scalabilité entre différents moteurs de base de données (SQLite, MySQL, PostgreSQL).

Conclusion

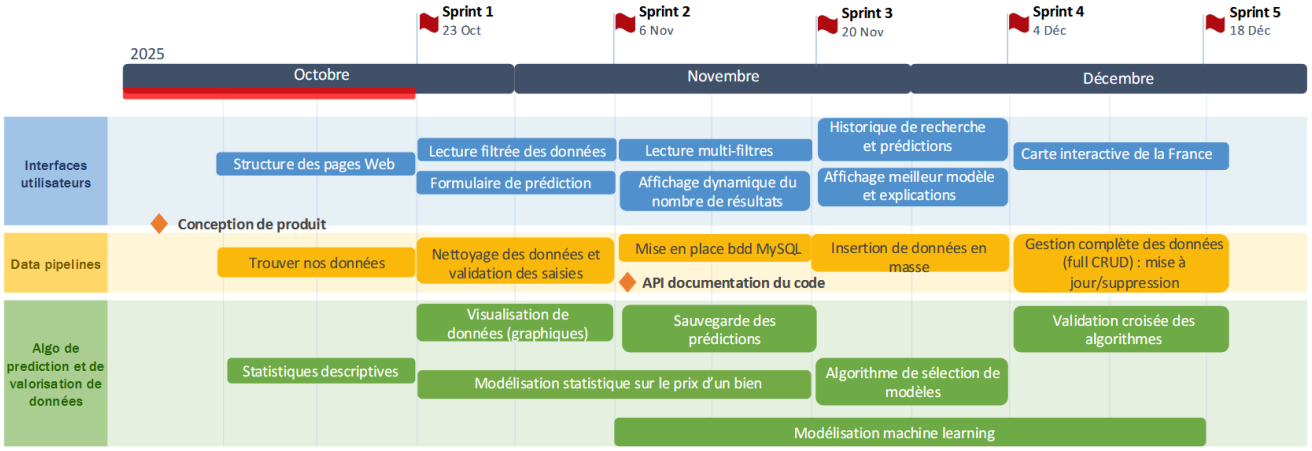
FPI est une application disponible via Internet sans installation, qui ne souffre pas de problèmes de performance et peut en réalité fonctionner pour des données sur une plus grande zone géographique, à condition d'ajuster les modèles pour les nouvelles données.
Elle souffre cependant de données pas assez détaillées pour de vraies études immobilières (adresses précises, proximité aux points d'intérêt, classe énergétique etc.) ce qui va à l'encontre du besoin initial identifié, que ce soit du point de vue acheteur ou vendeur.

Bibliographie / Références / Liens

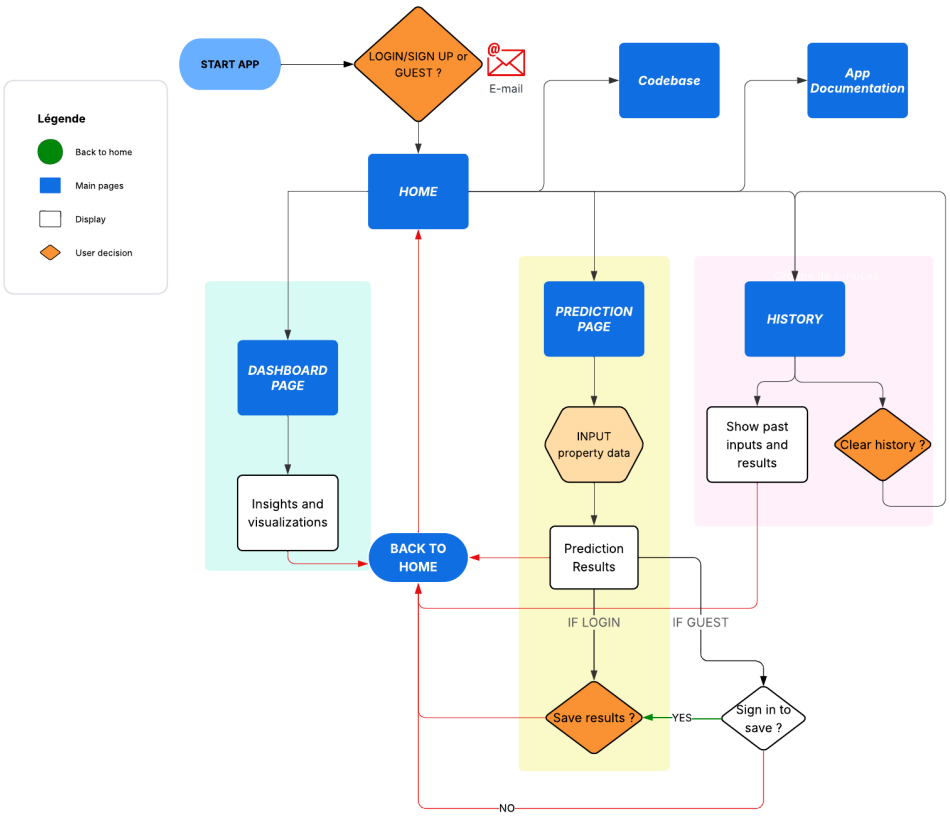
1. Le site
   https://gpd-m2sep-france-property-insight.onrender.com/
2. Lien GitLab
   https://gitlab-mi.univ-reims.fr/phan0005/gpd-m2sep-france-property-insight
3. Jeu de données
   https://www.data.gouv.fr/datasets/demandes-de-valeurs-foncieres/
4. Étude du contexte immobilier
   https://edito.meilleursagents.com/locaux-pros/points-marche/marche-immobilier
5. Outil pour l'interface utilisateur
   https://www.gradio.app/
6. Pour l'hébergement de l'application en ligne
   https://render.com/
7. Outil de documentation graphique
   https://www.mermaidchart.com/
8. Pour une API documentation du code
   https://pdoc.dev/
9. Logo FPI : France Property Insight
   https://www.vecteezy.com/vector-art/
10. Tufféry, S. (2017). *Data mining et statistique décisionnelle: L'intelligence des données*. Dunod.

Annexes

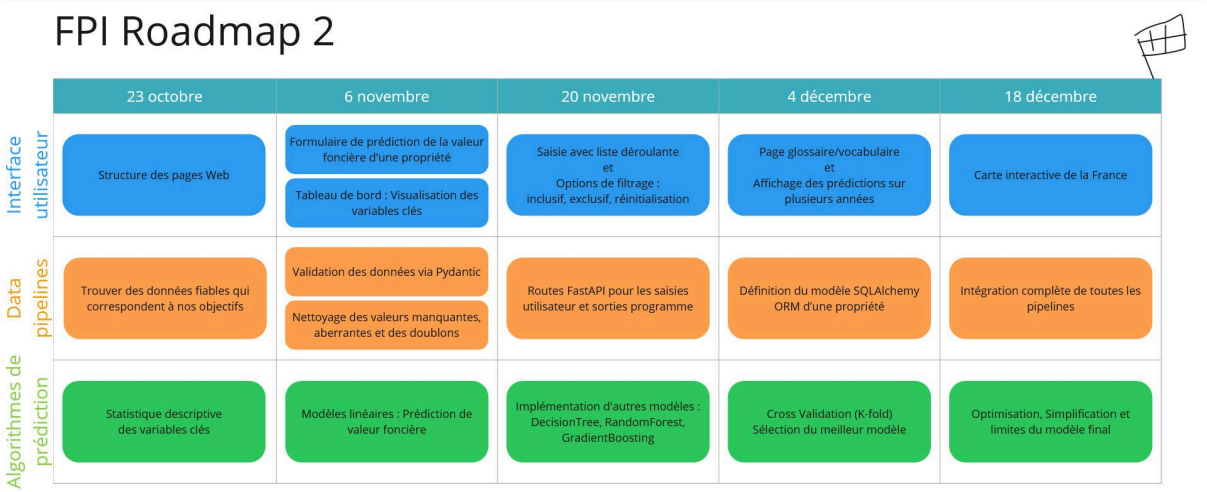
Roadmap V1 Retour Sprint 1



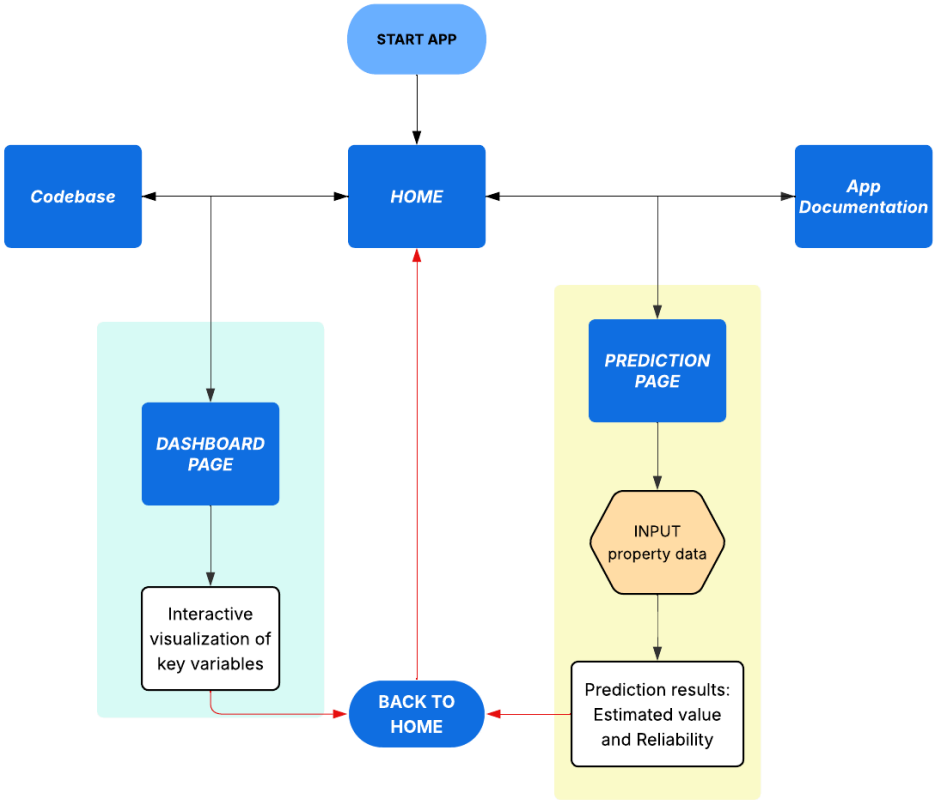
Logigramme V1



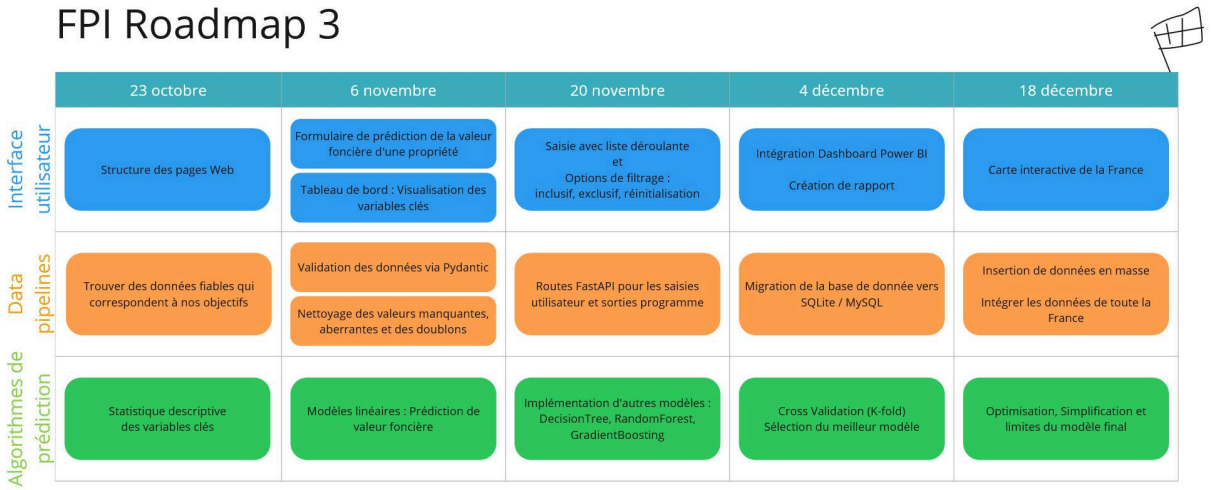
Roadmap V2



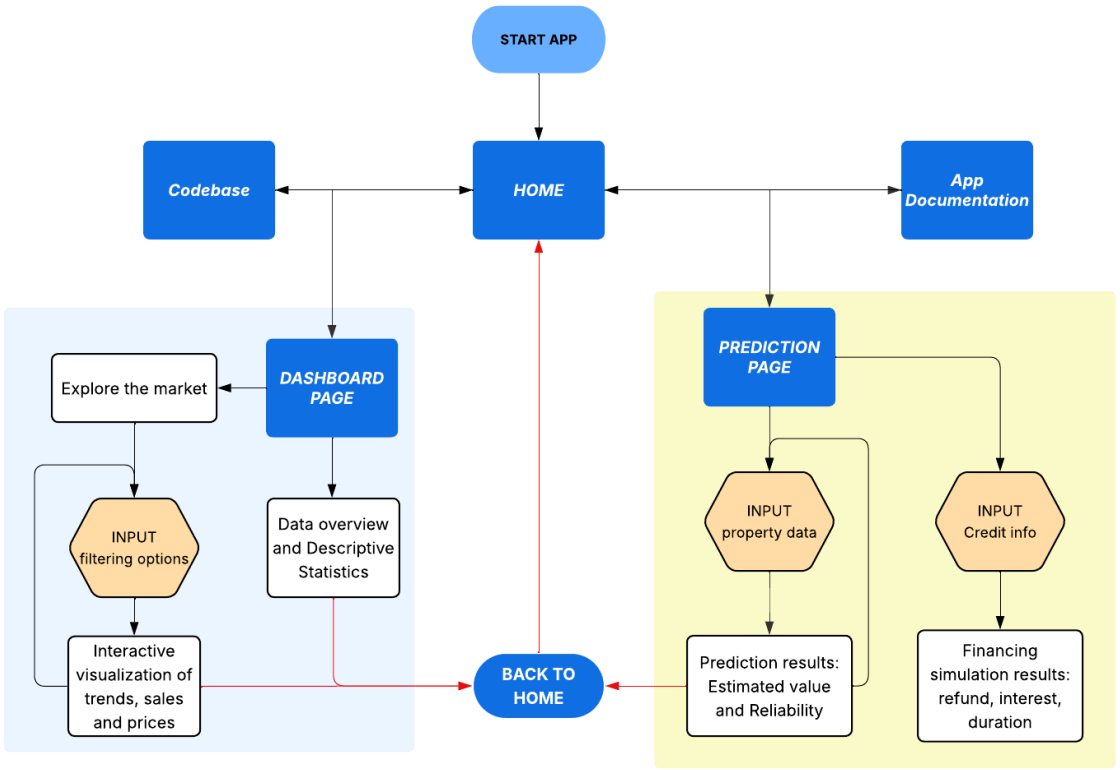
Logigramme V2



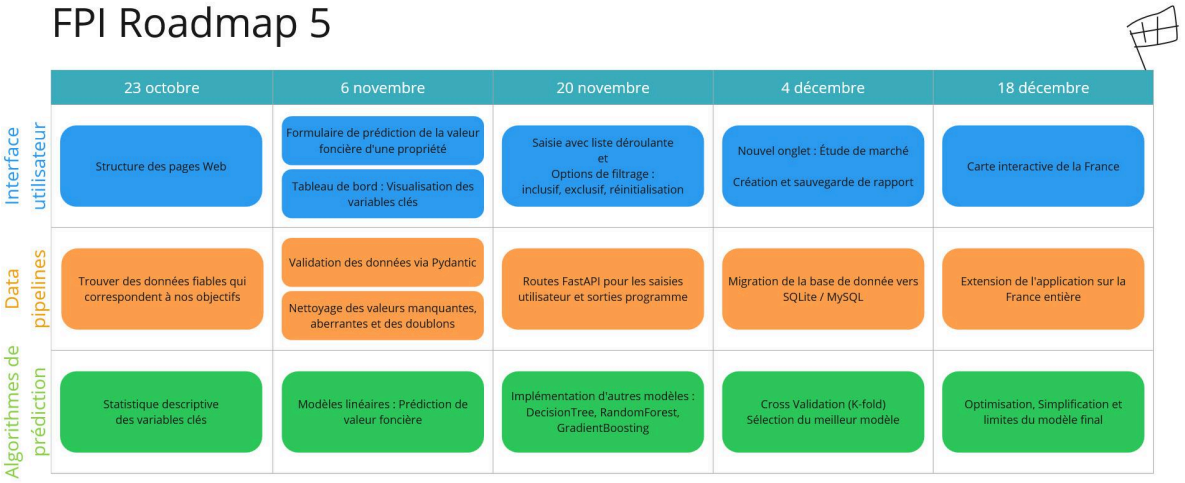
Roadmap V3



Logigramme V3



Roadmap V5



Logigramme V5